Stochastic Processes and their Applications 43 (1992) 191-222
North-Holland

Adaptive prediction and reverse martingales

Tomas Björk
Optimization and Systems Theory, Royal Institute of Technology, Stockholm, Sweden

Björn Johansson
Department of Mathematical Statistics, University of Stockholm, Stockholm, Sweden

Received 13 March 1991
Revised 18 September 1991

We study prediction problems for models where the underlying probability measure is not known. These problems are intimately connected with time reversal of Markov processes, and optimal predictors are shown to be characterized by being reverse martingales. For a class of diffusions we give a Feynman-Kac representation of the optimal predictor in terms of an associated complex valued diffusion and a concrete Wiener model is studied in detail. We also derive Cramér-Rao inequalities for the prediction error.

prediction * time reversal * martingales * diffusions * point processes * information inequalities

1. Introduction

In this paper we study adaptive prediction problems for a class of stochastic processes where the underlying probability measure is not known (it may, e.g., depend on a number of unknown parameters). A fairly common approach to problems of this kind consists in feeding a parameter estimate into a standard predictor derived for the case when the parameter is known, and while this method in many cases is perfectly sensible one may argue that, from a philosophical point of view, it is rather ad hoc.

An alternative approach is suggested in Johansson [10], where it is shown that much of the classical theory of unbiased parameter estimation can be transferred to a predictive setting. The main object of the present paper is to develop these ideas further and in particular to study the close connection which exists between unbiased prediction and time reversal of Markov proceesses.

In Sections 2-3 we present the necessary background material as well as some concrete models, and in Section 4 we study the 'Basic Equation' the solution of which is equivalent to a solution of our prediction problem. The basic equation is solved (partially) for two models and we point at the (sometimes) strange behaviour of the solutions.

Correspondence to: Dr. Tomas Björk, Optimization and Systems Theory, Royal Institute of Technology, S-10044 Stockholm, Sweden.

Section 5 is a digression on the connection between our problems and the theory of extremal families.

Section 6 is devoted to the connection between prediction and time reversal, and one of the central results is Theorem 6.2 which roughly says that the optimal predictor process is characterized by being a reverse martingale. In particular we show that the optimal predictor is the solution of an inverse boundary value problem for a certain backward operator.

The diffusion case is studied in Section 7 where the main result is Theorem 7.3 which, by introducing an associated complex-valued diffusion, gives a stochastic representation formula of Feynman-Kac type for the optimal predictor. Using this theory we then study a model of a Wiener process with unknown drift, for which we can give a fairly complete solution of the prediction problem in terms of necessary and sufficient conditions for existence as well as an analytical solution.

In Section 8 we discuss the particular problems which arise in connection with point process models, and discuss an approximately optimal predictor.

Sections 9–11 are concerned with information inequalities and we derive Cramér–Rao type results for the prediction error. In particular we derive analytic formulas for the predictive information matrix for a class of diffusions as well as for a class of point processes.

2. General background

Let $(\Omega, \mathcal{F})$ be a measurable space carrying the following objects:

(i) A stochastic process

$$X : [0, \infty) \times \Omega \to \mathbb{R}^k.$$

(ii) A family $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F}_\infty^X)$.

The intuitive interpretation of this setup is that the process X is governed by some $P \in \mathcal{P}$, where P is unknown to us. We are, however, allowed to observe X over time, and our object is to make 'good' predictions of X despite the fact that we do not know the correct P.

In order to make this problem a bit a more precise let us consider a fixed function

$$\Phi : \mathbb{R}^k \to \mathbb{R}, \tag{2.1}$$

and two points in time, t and T with $0 < t < T$.

The main problem. Find a 'good' predictor Z, of $\Phi(X_T)$, based on the information $\mathcal{F}_t^X$. Here 'good' is to be interpreted in terms of the mean-square loss function

$$C(Z, P) = E_P[\{(Z - \Phi(X_T)\}^2]. \tag{2.2}$$

Since the loss function C depends on P (which we do not know), the Main Problem is not a well posed optimization problem. However, in Johansson [10] it is shown that a large part of the classical theory of unbiased parameter estimation can be transferred to a predictive setting, thus enabling us to study *uniformly optimal* unbiased predictors. We will now briefly recapitulate some material from [10] which will be neeeded later on.

Assumption 2.1. For all P we have

$$E_P[\{\Phi(X_T)\}^2] < \infty. \tag{2.3}$$

Assumption 2.2. All $P \in \mathcal{P}$ are dominated by some base measure $P_0 \in \mathcal{P}$ on $\mathcal{F}_t^X$ for all $t \in [0, \infty)$.

Definition 2.3. A stochastic variable Z is called an $\mathcal{F}_t^X$-predictor if:

(i) $Z \in \mathcal{F}_t^X$.

(ii) $E_P[Z^2] < \infty$, for all $P \in \mathcal{P}$.

Definition 2.4. A predictor Z is said to be unbiased if

$$E_P[Z] = E_P[\Phi(X_T)], \quad \text{for all } P \in \mathcal{P}. \tag{2.4}$$

In order to have some data-reduction we will need a predictive version of statistical sufficiency (see, e.g., Takeushi and Akahira [14]).

Definition 2.5. An $\mathbb{R}^k$-valued process Y is said to be prediction sufficient for $(\Omega, \mathcal{F}, \mathcal{P}, X)$ if:

(i) Y is $\mathcal{F}^X$-adapted.

(ii) For each $t \in [0, \infty)$, Y_t is sufficient for $\mathcal{P}$ restricted to $\mathcal{F}_t^X$, i.e., for each bounded stochastic variable Z in $\mathcal{F}_t^X$, there exists a Borel-function f on $\mathbb{R}^k$ such that

$$E_P[Z \mid Y_t] = f(Y_t), \quad \text{for all } P \in \mathcal{P}. \tag{2.5}$$

(iii) For each t, the sigma-algebras $\sigma\{X_s;\ s \leq t\}$ and $\sigma\{X_s;\ s \geq t\}$ are conditionally independent given Y_t.

A prediction sufficient process Y thus contains all information in $\mathcal{F}_t^X$ relevant for identification of P, as well as for prediction purposes. The condition (iii) above is of course closely related to the Markov property, and we see that if a process Y satisfies the conditions

(a) for each t, $X_t \in \sigma\{Y_t\}$;

(b) Y is a $(P, \mathcal{F}^X)$-Markov process for each $P \in \mathcal{P}$;

then condition (iii) will be satisfied. This will in fact be the typical situation in the rest of this paper.

Definition 2.6. An $\mathbb{R}^k$-valued process Y is said to be complete for $\mathcal{P}$ if, for every fixed Borel-function g, the condition

$$E_P[g(Y_t)] = 0, \quad \text{for all } P \in \mathcal{P},$$

implies

$$g(Y_t) = 0, \quad \mathcal{P}\text{-a.s.}$$

(Sometimes we express this by simply saying that the model is complete.)

Now it is easy to transfer the theorems of Rao-Blackwell and Lehmann-Scheffé to the prediction setting [10].

Theorem 2.7 (Rao-Blackwell). *Suppose that Y is prediction sufficient and suppose that Z is an $\mathcal{F}_t^X$-predictor of $\Phi(X_T)$. Define the predictor $f(Y_t)$ by*

$$f(Y_t) = E_P[Z \mid Y_t]. \tag{2.6}$$

(Note that by predictive sufficiency the definition does not depend on P.) Then $f(Y_t)$ is uniformly better than Z in the sense that

$$E_P[\{f(Y_t) - \Phi(X_T)\}^2] \leq E_P[\{Z - \Phi(X_T)\}^2], \quad \textit{for all } P \in \mathcal{P}. \qquad \square \tag{2.7}$$

Theorem 2.8 (Lehmann-Scheffé). *Assume that Y is prediction sufficient and complete. Assume furthermore that there exists some unbiased $\mathcal{F}_t^X$-predictor Z of $\Phi(X_T)$. Define $f(Y_t)$ as in Theorem* 2.7. *Then $f(Y_t)$ is uniformly optimal in the class of unbiased predictors in the sense that* (2.7) *above holds for all P and for all unbiased $\mathcal{F}_t^X$-predictors Z of $\Phi(X_T)$. Furthermore $f(Y_t)$ is unique $\mathcal{P}$-a.s.* $\square$

Definition 2.9. An unbiased predictor V is called an UMSEUP ('Uniformly Minimum Squared Error Unbiased Predictor') if

$$E_P[\{V - \Phi(X_T)\}^2] \leq E_P[\{Z - \Phi(X_T)\}^2], \tag{2.8}$$

for all unbiased $\mathcal{F}_t^X$-predictors Z of $\Phi(X_T)$ and for all $P \in \mathcal{P}$.

3. Finding a prediction sufficient process

Given the model $(\Omega, \mathcal{F}, \mathcal{P}, X)$ the problem arises how to find a prediction sufficient process Y. One more or less obvious method when X is a Markov process is as follows.

(a) Compute the Radon-Nikodym derivative

$$L_t^P = \frac{dP}{dP_0}, \quad \text{restricted to } \mathcal{F}_t^X.$$

(This is typically done by applying the Girsanov theorem.)

(b) Use the factorization theorem (Lehmann [13, p. 55]) to obtain a sufficient process Y^0.

(c) In some cases Y^0 itself will be a prediction sufficient process. Generally speaking we cannot hope for this, but for a large class of models Y^0 will be an additive functional on X, so if we define Y by

$$Y = \begin{pmatrix} X \\ Y^0 \end{pmatrix},$$

this Y will satisfy the conditions (a) and (b) of the comment following Definition 2.5 and thus Y will be prediction sufficient.

To see how this works let us look at some concrete examples.

Example 3.1. *A Wiener process with unknown drift.* This will be one of our standard examples in the sequel. Let Ω be the space $C[0,\infty)$ and let X be the coordinate process on Ω. The family $\mathcal{P}$ is given by $\mathcal{P} = \{P_a \mid a \in \mathbb{R}\}$ where X under P_a satisfies

$$\begin{aligned} &dX_t = a\,dt + dW_t, \\ &X_0 = 0, \end{aligned} \tag{3.1}$$

where W is a standard Wiener process. The natural base measure is the Wiener measure P_0, and from Girsanov's theorem we have

$$L_t^a = e^{aX_t - a^2 t/2}. \tag{3.2}$$

It follows from the factorization theorem that we can choose X itself as a prediction sufficient process. This model is easily seen to be complete.

Example 3.2. *A diffusion with unknown drift.* Let Ω be the space $C[0,\infty)$ and let X be the coordinate process on Ω. Define P_0 as Wiener measure and consider as given a function

$$g : \mathbb{R} \to \mathbb{R},$$

such that the Novikov condition

$$E_0\left[\exp\left\{\tfrac{1}{2}a^2 \int_0^t g^2(X_s)\,ds\right\}\right] < \infty \tag{3.3}$$

holds for all $t > 0$ and for all a in $\mathbb{R}$. Now we define P_a on $\mathcal{F}_t^X$ by the formula

$$dP_a = L_t^a\,dP_0, \tag{3.4}$$

where

$$L_t^a = \exp\left\{a\int_0^t g(X_s)\,dX_s - \tfrac{1}{2}a^2 \int_0^t g^2(X_s)\,ds\right\}. \tag{3.5}$$

From Girsanov's theorem it now follows that, given P_a, X will be a (weak) solution of the stochastic differential equation

$$\begin{aligned} &dX_t = ag(X_t)\,dt + dW_t, \\ &X_0 = 0, \end{aligned} \tag{3.6}$$

and we define $\mathcal{P}$ by $\mathcal{P} = \{P_a \mid a \in \mathbb{R}\}$. We see from (3.5) that a sufficient process is given by the 2-dimensional process

$$Y_t^0 = \left(\int_0^t g(X_s)\, dX_s, \int_0^t g^2(X_s)\, ds \right).$$

Y^0 itself is generally not Markovian, so it is not a *prediction* sufficient process.

If, however, we add X as a component we see that a 3-dimensional prediction sufficient process is given by

$$\begin{aligned} Y_{1,t} &= X_t, \\ Y_{2,t} &= \int_0^t g(X_s)\, dX_s, \\ Y_{3,t} &= \int_0^t g^2(X_s)\, ds. \end{aligned} \tag{3.7}$$

For $g \equiv 1$, (3.7) collapses to Example 3.1, and for $g(x) = x$ we have the Ornstein-Uhlenbeck model

$$\begin{aligned} dX_t &= aX_t + dW_t, \\ X_0 &= 0. \end{aligned} \tag{3.8}$$

For this model we can use Ito's formula on the stochastic integral in (3.7) to see that a 2-dimensional prediction sufficient process is given by

$$\begin{aligned} Y_{1,t} &= X_t, \\ Y_{2,t} &= \int_0^t X_s^2\, ds. \end{aligned} \tag{3.9}$$

Example 3.3. *Geometric Brownian motion.* Using the same machinery as above we consider a model where, given P_a, the process X is the weak solution of

$$\begin{aligned} dX_t &= aX_t\, dt + X_t\, dW_t, \\ X_0 &= 1. \end{aligned} \tag{3.10}$$

The natural base measure is P_0, and it is easily seen that for all real a we have

$$L_t^a = (X_t)^a \cdot \exp\{\tfrac{1}{2}at(1-a)\}, \tag{3.11}$$

so we see that X itself is prediction sufficient.

This type of process is of particular interest in economic theory, where it is used to model randomly varying prices on e.g., the stock-market. We recall the well known fact that X is the exponential of a Wiener process with drift $a - \frac{1}{2}$.

Example 3.4. *A Poisson process with unknown (but constant) intensity.* This case has been studied in detail in [9]. Let Ω be the space of counting process trajectories and let N be the coordinate process on Ω (for mnemotechnic reasons we will use N instead of X). Let P_a be the measure on $(\Omega, \mathcal{F}_\infty^N)$ under which N is a Poisson process with intensity a for $a>0$, and define $\mathcal{P}$ by $\mathcal{P}=\{P_a \mid a>0\}$. All measures P_a are equivalent, and the natural base measure is the Poisson measure P_1. From Girsanov's theorem we have

$$L_t^a = a^{N_t}\, e^{(1-a)t}, \tag{3.12}$$

so from the factorization theorem we see that N itself is sufficient. Furthermore, since N is a Markov process for all P_a, we see that as a prediction sufficient process we can choose N itself. It is easily seen that this model is complete.

Example 3.5. *The Yule model.* This model has been studied in [9]. We let Ω and N be as in Example 3.4 but we let P_a be the measure under which N has the $\mathcal{F}_t^N$-intensity $\lambda_t = a(N_{t-}+1)$. From Girsanov's theorem we have

$$L_t^a = \exp\left\{\int_0^t \log\{a(N_{s-}+1)\}\, dN_s - \int_0^t (aN_{s-}+a-1)\, ds\right\}, \tag{3.13}$$

and we see that a prediction sufficient process is given by

$$\begin{aligned} Y_{1,t} &= N_t, \\ Y_{2,t} &= N_t t - \int_0^t N_s\, ds. \end{aligned} \tag{3.14}$$

The reason for this particular choice of Y instead of the more obvious pair consisting of

$$\begin{aligned} Y_{1,t} &= N_t, \\ Y_{2,t} &= \int_0^t N_s\, ds, \end{aligned} \tag{3.15}$$

is that Y in (3.14) grows by jumps only, which makes it easier to handle. This model is *not* complete (see Johansson [9]).

4. The basic equation

In this section we will give the basic equation which characterizes the UMSEUP (if and when it exists). We will partially solve the equation for the Poisson and Wiener models and point at the sometimes strange behaviour of the solutions. The deeper reason for this behaviour is explained by the results of Sections 6–7 (see Remarks 6.6 and 7.2). From the previous section we see that in all cases of interest (to us) the prediction sufficient statistic Y will contain the process X as a component. It is therefore natural to extend the problem of predicting X_T to that of predicting Y_T. To economize notation we therefore make the following assumption.

Assumption 4.1. The process X is prediction sufficient.

Now we turn back to our main poroblem of predicting $\Phi(X_T)$ based on $\mathcal{F}_t^X$, and because of Assumption 4.1 we need only to look at predictors of the form $f(t, X_t)$. The following (surprisingly easy) theorem characteizes the optimal f.

Theorem 4.2. *Suppose that X is prediction sufficient. Consider, for fixed t, T and Φ the equation*

$$E_P[f(t, X_t) | X_T = x] = \Phi(x), \quad \mathcal{P} \circ X_T^{-1}\text{-a.s.} \tag{4.1}$$

Then we have the following result.

(i) *If f solves (4.1) then $f(t, X_t)$ is an unbiased $\mathcal{F}_t^X$-predictor of $\Phi(X_T)$.*

(ii) *Assume that the model is complete, and that there exists some unbiased $\mathcal{F}_t^X$-predictor. Then (4.1) has a unique solution f, and $f(t, X_t)$ is the $\mathcal{P} \circ X_T^{-1}$-a.s. unique UMSEUP for $\Phi(X_T)$.*

Remark 4.3. Note that we can write E instead of E_P in (4.1) since Assumption 4.1 implies that the expected value does not depend on P.

Proof of Theorem 4.2. The proof of (i) is trivial so we turn to (ii): Suppose therefore that Z is an unbiased $\mathcal{F}_t^X$-predictor. Now we define f by

$$f(t, x) = E[Z | X_t = x], \tag{4.2}$$

where the expectation again does not depend on the choice of P. By completeness and Theorem 2.8 $f(t, X_t)$ will be unique $\mathcal{F}_t^X$-measurable UMSEUP of $\Phi(X_T)$. To see that f solves (4.1) we note that since $f(t, X_t)$ is an unbiased $\mathcal{F}_t^X$-predictor it will also be an unbiased $\mathcal{F}_T^X$-predictor of $\Phi(X_T)$ (but of course not an optimal one). It follows again from Theorem 2.8 that the expectation

$$E[f(t, X_t) | X_T] \tag{4.3}$$

will be the unique optimal $\mathcal{F}_T^X$-predictor of $\Phi(X_T)$. But the optimal $\mathcal{F}_T^X$-predictor of $\Phi(X_T)$ is obviously $\Phi(X_T)$ itself, so

$$E[f(t, X_t) | X_T] = \Phi(X_T), \quad \mathcal{P}\text{-a.s.}, \tag{4.4}$$

from which (4.1) follows. □

The main tasks are now:

(a) To give conditions on t, T and Φ which will ensure the existence of a solution to (4.1).

(b) To actually construct the solution.

In order to get a feeling for what is going on, let us go back to Examples 3.1 and 3.4.

The Poisson model (Example 3.4). This example is discussed in Johansson [10] and we only give the results. It is easily seen that for the Poisson model the basic equation (4.1) reads

$$\sum_{k=0}^{n}\binom{n}{k}\left(\frac{t}{T}\right)^{k}\left(1-\frac{t}{T}\right)^{n-k} f(t,k)=\Phi(n), \quad n=0,1,2,\ldots. \tag{4.5}$$

It is shown in [10] that (4.5) has a unique solution for every $t<T$ and for every Φ. The solution is given by

$$f(t,n)=\sum_{k=0}^{n}\binom{n}{k}\left(\frac{T}{t}\right)^{k}\left(1-\frac{T}{t}\right)^{n-k}\Phi(k), \quad n=0,1,2,\ldots. \tag{4.6}$$

Remark 4.4. Though the Poisson process is extremely well behaved in the sense that the basic equation can always be solved, it also possesses some alarming features.

(i) The sum in (4.6) looks like a binomial expectation. Note, however, that $T/t>1$, so the sum has alternating coefficients thus indicating numerical instability with respect to errors in Φ.

(ii) For some choices of Φ the formula (4.6) will produce a perfectly sensible result. If, e.g., $\Phi(n)=n$ we obtain the estimator

$$f(t,N_t)=N_t+(N_t/t)(T-t), \tag{4.7}$$

which is what you get if you compute the super-optimal predictor

$$E_a[N_T\,|\,N_t]=N_t+a(T-t), \tag{4.8}$$

and then plug in the maximum-likelihood estimator

$$\hat{a}=N_t/t. \tag{4.9}$$

For more 'irregular' choices of Φ you may, however, end up with a nonsensical estimator. Suppose, e.g., that $t=1$, $T=3$ and $\Phi(n)=I(n=0)$, where I denotes the indicator function. We get

$$f(1,n)=(-2)^n, \quad n=0,1,2,\ldots, \tag{4.10}$$

which is quite ridiculous since, for any $n\geq 1$, if you observe $N_1=n$ then you know that with probability one $N_3\geq n\geq 1$, so that $\Phi(N_3)=0$, $\mathcal{P}$-a.s. Furthermore we see that the predictor (4.10) is highly irregular in the sense that it fluctuates wildly. Similar disturbing phenomena will occur for every Φ with finite support, and problems of this type are in fact unavoidable for the kind of prediction problems treated here (see also remark 4.7). A fairly satisfying explanation is given in Remark 7.2.

The Wiener model (Example 3.1). In order to write down equation (4.1) in this case we have to compute the conditional density $q(t,x;T,y)$ of X_t given $X_T=y$, and since this distribution does not depend on a (by predictive sufficiency) we may as

well consider the simplest case, i.e., $a = 0$. Then we have the standard Wiener bridge, so

$$q(t, y; T, x) = K(t, T) \exp\left[-\frac{1}{2}\left\{ \frac{(x-y)^2}{T-t} + \frac{y^2}{t} - \frac{x^2}{T} \right\} \right], \tag{4.11}$$

where

$$K(t, T) = \sqrt{\frac{T}{2\pi t(T-t)}}. \tag{4.12}$$

After some rearrangement the basic equation becomes

$$K \int_{-\infty}^{+\infty} f(t, y) \exp\left\{ -\frac{1}{2}\frac{(x-y)^2}{T-t} \right\} \exp\left\{ -\frac{y^2}{2t} \right\} \mathrm{d}y = \exp\left\{ -\frac{x^2}{2T} \right\} \Phi(x). \tag{4.13}$$

Taking the Fourier transform of (4.13) with kernel $(1/\sqrt{2\pi}) \exp(-\mathrm{i}x\omega)$ gives us

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t, y)\varphi_t(y) \exp\{-\mathrm{i}\omega y\} \exp\{-\tfrac{1}{2}(T-t)\omega^2\} \, \mathrm{d}y \\ &\quad = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \varphi_T(x)\Phi(x) \exp\{-\mathrm{i}\omega x\} \, \mathrm{d}x, \end{aligned} \tag{4.14}$$

where $\varphi_t(\cdot)$ is the Gaussian density

$$\varphi_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\left\{ -\frac{x^2}{2t} \right\}. \tag{4.15}$$

Defining $H_{t,s}(\omega)$ by

$$H_{t,s}(\omega) = \exp\{\tfrac{1}{2}(t-s)\omega^2\}, \tag{4.16}$$

and denoting the Fourier transform by $\mathcal{G}$ (in order to avoid confusion with the various sigma-algebras) (4.14) now reads as

$$\mathcal{G}\{f \cdot \varphi_t\} = H \cdot \mathcal{G}\{\Phi \cdot \varphi_T\}. \tag{4.17}$$

We thus have the following result.

Proposition 4.5. *Consider the model* (3.2). *Suppose that* t, T *and* Φ *are such that*

(i) $\Phi \cdot \varphi_t \in L^1(\mathbb{R}, \mathrm{d}x)$,

(ii) $H_{T,t} \cdot \mathcal{G}\{\Phi \cdot \varphi_T\} \in L^1(\mathbb{R}, \mathrm{d}\omega)$.

Then a unique solution of (4.1) *exists and is given by*

$$f(t, x) = \varphi_t(x)^{-1} \mathcal{G}^{-1}\{H_{T,t} \cdot \mathcal{G}[\Phi \cdot \varphi_T]\}(x). \qquad \square \tag{4.18}$$

Given enough integrability we get a slightly neater formula, using $\hat{}$ to denote the Fourier transform.

Corollary 4.6. *Suppose that* $\Phi, \hat{\Phi} \in L^1$ *and that* (i)–(ii) *of Proposition* 4.5 *are satisfied. Then the unique solution of* (4.1) *is given by*

$$f(t, x) = \mathcal{G}^{-1}\{H_{T^2/t,T} \cdot \hat{\Phi}\}(xT/t). \tag{4.19}$$

Proof. We have

$$\begin{aligned} H_{T,t}(\omega)\mathcal{G}\{\Phi \cdot \varphi_T\}\{\omega\} &= H_{T,t}(\omega)(\hat{\Phi} * \hat{\varphi}_t)(\omega) \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \exp\{\tfrac{1}{2}[(T-t)\omega^2 - T(\omega-\lambda)^2]\}\hat{\Phi}(\lambda)\,d\lambda. \end{aligned} \tag{4.20}$$

Completing the square for ω, the exponent becomes

$$-\tfrac{1}{2}t\left[\omega - \frac{T\lambda}{t}\right]^2 + \tfrac{1}{2}\lambda^2\left[\frac{T^2}{t} - T\right], \tag{4.21}$$

so, using (4.18) and applying the Fubini theorem when taking the inverse transform, we obtain

$$f(t, x) = \frac{1}{2\pi}(\varphi_t(x))^{-1}\int_{-\infty}^{+\infty} \exp\left\{i\frac{xT\lambda}{t}\right\}\varphi_t(x)\exp\left\{\tfrac{1}{2}\lambda^2\left[\frac{T^2}{t} - T\right]\right\}\hat{\Phi}(\lambda)\,d\lambda$$

which gives us (4.19). □

Remark 4.7. The most disturbing fact about (4.18) and (4.19) is the appearance of the factor H which is a *positive* exponential. This has two consequences (c.f. Remark 4.4).

(i) H amplifies high frequencies thus indicating instability of the solution with respect to small changes in Φ. See Remark 7.2.

(ii) If, e.g., $\hat{\Phi}$ is to be in L^1 then $\hat{\Phi}$ must be very rapidly decreasing. Thus Φ itself is not allowed to be too rapidly decreasing. Consider as an example,

$$\Phi(x) = \exp\{-c(\tfrac{1}{2}x^2)\} \quad (c > 0). \tag{4.22}$$

Using Corollary 4.6 it is easy to see that a solution exists if

$$c < t/(T(T-t)). \tag{4.23}$$

Thus, for fixed t and T, c is not allowed to be to big. For fixed c, on the other hand, we see from (4.23) that there always exists an UMSEUP for this Φ provided t and T are close enough. This reflects the intuitively obvious fact that it is easier to predict the near future than it is to make predictions far ahead.

If we choose $\Phi(x) = x$ we obtain, as in the Poisson case, the natural predictor

$$f(t, X_t) = X_t + (X_t/t)(T-t). \tag{4.24}$$

5. Maximal families

This section is a slight digression into the theory of maximal and extremal families. The main point to be made is that if we have managed to solve the basic equation (4.1) for a fixed model, then in many cases we have in fact also solved it for a much more complicated model. The arguments are sometimes rather informal and the reader is referred to Lauritzen [12] for the technical details.

Let us again consider a fixed model $(\Omega, \mathcal{F}, \mathcal{P}, X)$ where X is assumed to be prediction sufficient.

Definition 5.1. For any probability measure P on $(\Omega, \mathcal{F}_\infty^X)$ we define $P^{t,x}$ on $(\Omega, \mathcal{F}_t^X)$ by

$$P^{t,x}(A) = P(A \mid X_t = x), \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}^k. \tag{5.1}$$

We note that, because of predictive sufficiency, there exists a fixed family of probability measures $Q_{t,x}$, indexed by (t, x), such that

$$Q_{t,x} = P^{t,x} \quad \text{for every } P \in \mathcal{P}, \text{ and for all } (t, x) \in \mathbb{R}_+ \times \mathbb{R}^k. \tag{5.2}$$

We also note that the expectation operator entering the basic equation (4.1) is really $Q_{T,x}$ (rather than the original P). Thus: if we have solved (4.1) for a particular model, then we have actually solved it for every probability measure P such that

$$Q_{t,x} = P^{t,x} \quad \text{for all } (t, x) \in \mathbb{R}_+ \times \mathbb{R}^k.$$

We are thus led to the following definition.

Definition 5.2. The maximal family $\mathcal{M}$ generated by $\mathcal{P}$ is the class of all probability measures P on $(\Omega, \mathcal{F}_\infty^X)$ such that

$$Q_{t,x} = P^{t,x} \quad \text{for all } (t, x) \in \mathbb{R}_+ \times \mathbb{R}^k,$$

where Q is given by (5.2).

It is obvious that $\mathcal{M}$ is a convex set, and we denote its set of extremal points by $\mathcal{E}$. Given sufficient regularity it can be shown (Lauritzen [12, p. 196, Proposition IV.I.I]) that the conclusion of the Krein-Milman theorem holds for $\mathcal{M}$, i.e., for each $P \in \mathcal{M}$ there exists a probability measure v on $\mathcal{E}$ such that

$$P = \int_{\mathcal{E}} P_e \, dv(e). \tag{5.3}$$

In other words: every element in $\mathcal{M}$ is a mixture of the extremal elements P_e. Let us now take a look at some of our examples in Section 3 for this point of view.

Example 5.3. *The Wiener model.* We define the family $\mathscr{P}$ as in Example 3.1. In this case it can be shown that the extremal family $\mathscr{E}$ is given by $\mathscr{P}$ itself. The maximal family $\mathscr{M}$ is thus the set of mictures of elements of $\mathscr{P}$, and in view of (5.3) this means that for every P in $\mathscr{M}$ there exists a probability measure v on the real line, such that

$$P = \int_{-\infty}^{+\infty} P_a \, dv(a). \tag{5.4}$$

In more concrete terms (5.4) says that, under P, the process X satisfies the stochastic differential equation

$$\begin{aligned} dX_t &= Z(\omega)\, dt + dW_t, \\ X_0 &= 0, \end{aligned} \tag{5.5}$$

where Z is a stochastic variable with distribution v, which is independent of W. Thus the optimal predictor for the model (3.1), derived in Theorem 4.5, is in fact optimal for the entire model (5.5).

It is not a priori evident that the process X defined by (5.5) is a Markovian diffusion process with respect to the filtration $\mathscr{F}^X$, but it is fairly easy to show that X in fact satisfies

$$dX_t = \mu(t, X_t)\, dt + dW_t, \tag{5.6}$$

where μ is given by

$$\mu(t, x) = \frac{\int_{\mathbb{R}} z \exp\{zx - \frac{1}{2}z^2 t\}\, dv(z)}{\int_{\mathbb{R}} \exp\{zx - \frac{1}{2}z^2 t\}\, dv(z)}. \tag{5.7}$$

By choosing v as a Gaussian distribution we see that, e.g., all models of the type (5.6) with μ of the form

$$\mu(t, x) = \frac{x + \alpha}{t + \beta}, \quad \alpha \in \mathbb{R}, \ \beta > 0, \tag{5.8}$$

are included in the model (5.5).

Example 5.4. *The Poisson model.* We define $\mathscr{P}$ as in Example 3.4. This case is more or less parallel to the Wiener model. The extremal family $\mathscr{E}$ is given by $\mathscr{P}$ itself, and a typical member P of the maximal family $\mathscr{M}$ is a mixture of Poisson measures

$$P = \int_0^{\infty} P_a \, dv(a), \tag{5.9}$$

where v is a probability measure on the positive reals. In process terms this means that, under P, the counting process N is a Cox process with $\mathscr{F}_t^N \vee \sigma(Z)$-intensity

$$\lambda_t = Z,$$

where Z is a stochastic variable with distribution v (in this case N is commonly referred to as a weighted Poisson process). The predictable $\mathcal{F}_t^N$-intensity is easily seen to be given by

$$\lambda_t = \frac{\int_0^\infty z^{N_{t-}+1} e^{-zt}\, dv(z)}{\int_0^\infty z^{N_{t-}} e^{-zt}\, dv(z)}, \tag{5.10}$$

(see, e.g., Bremaud [3, p. 173]), and if we choose a $\Gamma(\alpha^{-1}, \alpha\beta)$-distribution for Z we see that the Polya model

$$\lambda_t = \alpha \frac{1+\beta N_{t-}}{1+\alpha N_{t-}}, \quad \alpha > 0,\ \beta > 0, \tag{5.11}$$

is included in the maximal family.

Example 5.5. *The Yule model.* We consider the Yule model discussed in Example 3.5. It can be shown (Jensen and Johansson [8]) that the extremal family generated by $\mathcal{P}$ is given by

$$\mathcal{E} = \{P_{\alpha,\beta} \mid \alpha > 0, \beta > 0\}, \tag{5.12}$$

where N has the intensity

$$\lambda_t = \alpha \exp(\beta t)$$

under the measure $P_{\alpha,\beta}$. As noted in Example 3.5 the Yule model is not complete. The model generated by the maximal family $\mathcal{M}$ is, however, complete.

A study of the basic equation for the model generated by the maximal family is given in Johansson [9]. It was seen that, due to irregularities of the conditional distributions, an UMSEUP does not generally exist. Even the simple case of predicting the value of N_T does not admit an unbiased predictor. However, the following predictor $f(t, N_t, Y_t)$ was suggested as coming very close to being unbiased (in order to have a more readable notation we write (N, Y) instead of the pair (Y_1, Y_2) in (3.14)).

$$\begin{aligned} &f(t,0,0) = 0, \qquad f(t,1,x) = \Phi(1,x)\frac{1}{t}, \\ &f(t,m,x) = h(m,x)\left\{\frac{1}{t} g_m\left(\frac{x}{t}\right)\right\}^{-1}, \quad m \geq 2, \end{aligned} \tag{5.13}$$

where

$$\begin{aligned} h(m,x) = \sum_{n=1}^{m-1} \Bigg\{ &\binom{m}{n}\left(\frac{T}{t}\right)^n \left(1-\frac{T}{t}\right)^{m-n} \\ &\times \int \frac{\Phi(n,y)}{T(T-t)} g_{m-n}\left(\frac{y-x-(m-n)s}{T-t}\right) g_n\left(\frac{y}{T}\right) dy \Bigg\}, \end{aligned} \tag{5.14}$$

and g_n denotes the density of a sum of n independence random variables, each having a uniform distribution on the unit interval.

The nonexistence of a solution to the basic equation will be discussed from a different point of view in Section 8, where the predictor (5.13) will turn up again.

6. The martingale characterization

Up to this point we have tried to solve (4.1) for t and T fixed. In this section we will keep T fixed while allowing t to vary, and we will study properties of the associated process $f(t, X_t)$. As before X is a prediction sufficient process for the model.

Definition 6.1. The decreasing family of sigma-algebras $\mathcal{G}^X$ is defined by

$$\mathcal{G}_t^X = \sigma\{X_s;\ s \geq t\}. \tag{6.1}$$

We can now state one of the central results of this paper.

Theorem 6.2. (I) *Suppose that there exists a time S with $0 < S < T$, and a function*

$$f: [S, T] \times \mathbb{R}^k \to \mathbb{R},$$

such that the process

$$M_t = f(t, X_t) \tag{6.2}$$

has the following properties:

(i) *M is a reverse $(P, \mathcal{G}^X)$-martingale on $[S, T]$ for all $P \in \mathcal{P}$.*

(ii) *$M_T = \Phi(X_T)$, $\mathcal{P}$-a.s.*

Then, for each $t \in [S, T]$, $f(t, X_t)$ is an unbiased $\mathcal{F}_t^X$-predictor of $\Phi(X_T)$.

(II) *Suppose that the model is complete and suppose that for some fixed time $S < T$ there exists an unbiased $\mathcal{F}_S^X$-predictor. Then*:

(iii) *Equation (4.1) has a unique solution for every $t \in [S, T]$.*

(iv) *The process $M_t = f(t, X_t)$, is a reverse $(P, \mathcal{G}^X)$-martingale on $[S, T]$ for all $P \in \mathcal{P}$ with $M_T = \Phi(X_T)$, $\mathcal{P}$-a.s.*

Moreover: for each $t \in [S, T]$, $f(t, X_t)$ is the unique UMSEUP of $\Phi(X_T)$.

Proof. (I) Suppose that M given by (6.2) is a reverse martingale with $M_T = \Phi(X_T)$, $\mathcal{P}$-a.s. Then we immediately have

$$E_P[\Phi(X_T)] = E_P[M_T] = E_P[M_t] = E_P[f(t, X_t)]. \tag{6.3}$$

Thus $f(t, X_t)$ is an unbiased $\mathcal{F}_t^X$-predictor of $\Phi(X_T)$.

(II) Suppose that Z is an unbiased $\mathcal{F}_S^X$-predictor for some fixed $S < T$. For any t with $S \leq t \leq T$ we define M_t by

$$M_t = E_P[Z \mid X_t]. \tag{6.4}$$

Then we can write

$$M_t = f(t, X_t), \tag{6.5}$$

for some Borel-function f, and it follows from the Lehmann-Schaffé theorem that M_t is the unique UMSEUP for $\Phi(X_T)$ based on $\mathcal{F}_t^X$. For any $s<t$ we may now more or less repeat the argument in the proof of Theorem 4.2 to obtain

$$E_P[f(s, X_S)\,|\,X_t] = f(t, X_t), \tag{6.6}$$

and in particular we have

$$E_P[f(s, X_S)\,|\,X_T] = \Phi(X_T), \tag{6.7}$$

which shows that f solves equation (4.1) and that $M_T = \Phi(X_T)$, $\mathcal{P}$-a.s.

Since X is prediction sufficient it is a Markov process in forward time. Thus it is also a Markov process in backward time, so from (6.6) we get

$$E_P[f(s, X_S)\,|\,\mathcal{G}_t^X] = E_P[f(s, X_S)\,|\,X_t] = f(t, X_t),$$

which proves that M is a reverse martingale. □

The problem of finding an unbiased predictor process is thus more or less reduced to that of finding a reverse martingale with the correct boundary value, and in a Markovian context this is a standard task. As we noted in the proof above X will also be a Markov process in backward time and this fact will now be used.

Definition 6.3. The process $\tilde{X}$ defined on $[0, T]$ is given by

$$\tilde{X}_r(\omega) = X_{T-r}(\omega), \tag{6.8}$$

and the filtration $\tilde{\mathcal{F}}^X$ is defined by

$$\tilde{\mathcal{F}}_r^X = \sigma\{\tilde{X}_\tau;\, 0 \leq \tau \leq r\}. \tag{6.9}$$

For any function $f:[0, T]\times\mathbb{R}^k \to \mathbb{R}$ we define $\tilde{f}:[0, T]\times\mathbb{R}^k \to \mathbb{R}$ by

$$\tilde{f}(r, x) = f(T-r, x). \tag{6.10}$$

Generally speaking $\tilde{X}$ will be a much more unpleasant object that X. It will typically be time-dependent with singular behaviour near $r = T$. Note however that, since X is prediction sufficient, the transition probabilities of $\tilde{X}$ will be the same regardless of which measure $P \in \mathcal{P}$ we consider. In particular we can give the following definition independently of the choice of P.

Definition 6.4. Let $\tilde{A}(r)$ denote the infinitesimal operator of $\tilde{X}$ at time r, and let $\mathcal{D}[\tilde{A}(r)]$ denote its domain.

Now we have an immediate corollary of Theorem 6.2.

Corollary 6.5. *Supose that for some d in* $[0, T]$ *there exists a function*

$$f:(T-d, T]\times\mathbb{R}^k \to \mathbb{R}$$

such that

(i) *for each r in* $[0, d)$ *we have*

$$\tilde{f}(r, \cdot) \in \mathcal{D}[\tilde{A}(r)], \tag{6.11}$$

(ii) $\tilde{f}$ *solves the boundary value problem*

$$\begin{aligned} &\frac{\partial \tilde{f}}{\partial r}(r, x) + [\tilde{A}(r)\tilde{f}(r, \cdot)](x) = 0, \quad \text{on } [0, d) \times \mathbb{R}^k, \\ &\tilde{f}(0, x) = \Phi(x). \end{aligned} \tag{6.12}$$

Then, for all t in $(T-d, T]$, $f(X_t)$ *is an unbiased* $\mathcal{F}_t^X$*-predictor of* $\Phi(X_T)$. *If the model is complete the predictor is also the unique UMSEUP.*

Proof. Apply $\tilde{A}$ to $\tilde{X}$, use Dynkin's formula and Theorem 6.2. □

Remark 6.6. It is important to notice that the equation (6.12) is the *inverse* problem of an ordinary Kolmogorov-type backward equation (where the natural boundary conditions would be given at $r = T$). In other words: we are trying to invert the semigroup generated by $\tilde{X}$—a fact which can also be seen directly from equation (4.1).

It is sometimes convenient to write equation (6.12) in forward time t rather than reverse time r. We have

$$\begin{aligned} &\frac{\partial f}{\partial t}(t, x) = [\bar{A}(t)f(t, \cdot)](x), \quad \text{on } (T-d, T] \times \mathbb{R}^k, \\ &f(T, x) = \Phi(x), \end{aligned} \tag{6.13}$$

where

$$\bar{A}(t) = \tilde{A}(T-t).$$

Equations (6.13) and (6.12) are both 'backward' equations in the sense that the action of the operator is on the 'backward' variables t and X_t, while T is being held fixed. There is also a corresponding 'forward' equation where the action is on s and X_s where t is being held fixed and $s > t$.

Definition 6.7. Consider a stochastic variable Z such that

$$Z \in L^2(\Omega, \mathcal{F}_s^X, P), \quad \text{for all } P \in \mathcal{P}, \tag{6.14}$$

for some $s > t$.

If there exists a unique $\mathcal{F}_t^X$-measurable UMSEUP of Z we denote it by

$$\Pi_t[Z]. \tag{6.15}$$

Proposition 6.8. *Consider t and* Φ *as fixed, assume that the model is complete, and suppose that for some* $T > t$,

(i) $\Phi \in \mathcal{D}[\bar{A}(s)], \quad s \in [t, T]$,

(ii) $\Pi_t[\Phi(X_s)]$ *and* $\Pi_t[\{\bar{A}(s)\Phi\}(X_s)]$ *exists for all* $s \in [t, T]$.

Then we have the forward equation

$$\frac{\partial \Pi_t[\Phi(X_s)]}{\partial s} = -\Pi_t[\{\bar{A}(s)\Phi\}(X_s)], \quad s \in [t, T],$$
$$\Pi_t[\Phi(X_t)] = \Phi(X_t). \tag{6.16}$$

Proof. Applying Dynkin's formula to $\tilde{X}$ we have

$$\Phi(\tilde{X}_{T-t}) - \Phi(\tilde{X}_{T-s}) = \int_{T-s}^{T-t} [\tilde{A}(r)\Phi](\tilde{X}_r)\,\mathrm{d}r + \tilde{M}_{T-t} - \tilde{M}_{T-s}, \tag{6.17}$$

where $\tilde{M}$ is a $(\mathscr{P}, \tilde{\mathscr{F}}^X)$-martingale for all $P \in \mathscr{P}$. Going back to forward time in (6.17) and taking expectations we have for all $P \in \mathscr{P}$,

$$E_P[\Phi(X_s)] = E_P[\Phi(X_t)] - \int_t^s E_P[\{\bar{A}(u)\Phi\}(X_u)]\,\mathrm{d}u. \tag{6.18}$$

Now consider the stochastic variable $Z_{t,s}$ given by

$$Z_{t,s} = \Phi(X_t) - \int_t^s \Pi_t[\{\bar{A}(u)\Phi\}(X_u)]\,\mathrm{d}u. \tag{6.19}$$

and observe that $Z_{t,s}$ is $\mathscr{F}_t^X$-measurable. Taking expectations in (6.19) and using (6.18) we have

$$E_P[Z_{t,s}] = E_P[\Phi(X_S)], \quad \text{for all } P \in \mathscr{P}, \tag{6.20}$$

so $Z_{t,s}$ is an $\mathscr{F}_t^X$-predictor of $\Phi(X_s)$. Since the model is complete $Z_{t,s}$ is the unique UMSEUP of $\Phi(X_S)$, i.e.,

$$Z_{t,s} = \Pi_t[\Phi(X_s)]. \tag{6.21}$$

Combining (6.19) and (6.21) gives us (6.16). □

In a suggestive but somewhat careless notation we can denote optimal predictors by $\hat{}$ and write (6.16) as

$$\frac{\mathrm{d}\hat{\Phi}_s}{\mathrm{d}s} = -\widehat{[\bar{A}(s)\Phi]}_s,$$
$$\hat{\Phi}_t = \Phi. \tag{6.22}$$

The intuitive content of Proposition 6.8 can then be expressed by saying that prediction of the future of the process X is, in a sense, equivalent to the prediction of the future (in forward time) infinitesimal characteristics of the reversed process. If, in particular, X is a diffusion then the reversed process will also be a diffusion (under rather mild conditions, see Section 7 below), and if we take $\Phi(x) = x$ we see that the right-hand side of (6.22) will be a prediction of the future values of the drift coefficient of the reversed process.

7. Reversed diffusions

If the prediction sufficient process X is a diffusion process which, under the measure P satisfies the stochastic differential equation

$$\begin{aligned} dX_t &= \mu_P(t, X_t)\,dt + \sigma_P(t, X_t)\,dW_t, \\ X_0 &= x_0. \end{aligned} \tag{7.1}$$

Then, under some fairly mild conditions (see Haussmann and Pardoux [6]), the reversed process $\tilde{X}$ will also be a diffusion process of the form

$$\begin{aligned} d\tilde{X}_r &= \tilde{\mu}(r, \tilde{X}_r)\,dr + \tilde{\sigma}(r, \tilde{X}_r)\,d\tilde{W}_r, \\ \tilde{X}_0 &= X_T, \end{aligned} \tag{7.2}$$

where $\tilde{W}$ is a Wiener process in reverse time. We note again that, because of predictive sufficiency, the infinitesimal characteristics of the reversed process are the same for all $P \in \mathcal{P}$.

In this case the equation (6.12) is a boundary value problem for a partial differential equation:

$$\begin{aligned} &\frac{\partial \tilde{f}}{\partial r} + \sum_{i=1}^{k} \tilde{\mu}_i \frac{\partial \tilde{f}}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{k} \tilde{\alpha}_{i,j} \frac{\partial^2 \tilde{f}}{\partial x_i\, \partial x_j} = 0, \\ &\tilde{f}(0, x) = \Phi(x), \end{aligned} \tag{7.3}$$

where

$$\tilde{\alpha}_{i,j} = \sum_n \tilde{\sigma}_{i,n} \tilde{\sigma}_{j,n}.$$

The nature of (7.3) is perhaps seen more clearly if we write it in forward time.

Definition 7.1. Given (7.1), the function $\bar{\mu}$, $\bar{\sigma}$, $\bar{\alpha}$ are defined by

$$\begin{aligned} \bar{\mu}(t, x) &= \tilde{\mu}(T - t, x), \\ \bar{\sigma}(t, x) &= \tilde{\sigma}(T - t, x), \\ \bar{\alpha}(t, x) &= \tilde{\alpha}(T - t, x). \end{aligned} \tag{7.4}$$

In this notation (7.3) becomes

$$\begin{aligned} &\frac{\partial f}{\partial t} = \sum \bar{\mu}_i \frac{\partial f}{\partial x_i} + \frac{1}{2} \sum \bar{\alpha}_{i,j} \frac{\partial^2 f}{\partial x_i\, \partial x_j}, \\ &f(T, x) = \Phi(x), \end{aligned} \tag{7.5}$$

which is an *inverse* parabolic boundary value problem.

Remark 7.2. The strange behaviour of the optimal predictors (c.f. Remarks 4.4 and 4.7) is perhaps best understood by considering equation (7.5) above in the particular

case when X is a scalar process. Then we see that in (7.5) we are trying to solve (a version of) the heat equation backwards in time. Now, solving the heat equation in forward time has of course an extremely regularizing effect on boundary data (they will immediately be reproduced as entire analytic functions). Thus the process of solving the heat equation backwards will have a highly 'irregularizing' effect, and among other things be characterized by the following.

(i) A necessary condition for the existence of a solution is that the boundary data Φ is extremely regular (at least analytic).

(ii) The solution will be more irregular the further we get from the final time T.

(iii) The mapping from boundary data to $f(t, \cdot)$ for a fixed t will not be continuous in, e.g., supremum norm. In fact, the problem (7.5) is a typical extreme of an 'ill posed' problem.

Despite the problems discussed above we may, however, obtain a stochastic representation formula for the solution of (7.5) (when such a solution exists). First let us define $\bar{A}$ by

$$\bar{A} = \sum_i \bar{\mu}_i \frac{\partial}{\partial x_i} + \sum_{i,j} \bar{\alpha}_{i,j} \frac{\partial^2}{\partial x_i \, \partial x_j}. \tag{7.6}$$

Theorem 7.3. *Suppose that*:

(i) *The boundary function Φ can be extended to an entire analytic function on C^k.*

(ii) *There exists a time $d < T$ such that equation (7.5) has a solution $f(t, x)$ on $(d, T]x\mathbb{R}^k$ which, for all t in $(d, T]$, can be extended to an entire analytic function on C^k such that*

$$E_P[|f(t, X_t)|^2] < \infty, \quad \text{for all } P \in \mathcal{P}. \tag{7.7}$$

(iii) *for all t in $(d, T]$ the functions $\bar{\sigma}(t, \cdot)$ and $\bar{\mu}(t, \cdot)$ can be extended to entire analytic functions on C^k.*

(iv) *For each fixed (t, x) in $(d, T]x\mathbb{R}^k$ the complex-valued stochastic differential equation*

$$\begin{aligned} &dY_s = -\bar{\mu}(s, Y_s)\,ds + i\bar{\sigma}(s, Y_s)\,dW_s, \\ &Y_t = x, \end{aligned} \tag{7.8}$$

has a solution on $(d, T]$ satisfying the condition

$$E_{t,x}\left[\int_d^T \|(\nabla_x f)(s, Y_s)\bar{\sigma}(s, Y_s)\|^2\,ds\right] < \infty, \tag{7.9}$$

Then f has the stochastic representation

$$f(t, x) = E_{t,x}[\Phi(Y_T)], \tag{7.10}$$

where $E_{t,x}$ is the expectation operator induced by the condition $Y_t = x$.

Proof. Apply the Ito formula to $f(s, Y_s)$. □

Remark 7.4. We note that if P were known to us then the optimal $\mathcal{F}_t^X$-predictor of $\Phi(X_T)$ would obviously be $g(t, X_t)$, where

$$g(t, x) = E_{P,t,x}[\Phi(X_T)]. \tag{7.11}$$

Here $E_{P,t,x}$ is the expectation operator induced by the initial condition $X_t = x$ in (7.1). The problem with (7.11) is, of course, that in order to use the formula we have to know P, and the moral of Theorem 7.3 is that the optimal *adaptive* predictor *also* has the structure (7.11). The price we have to pay for not knowing P is the introduction of time-reversal and the complex-valued process Y.

Since the problem (7.5) is ill-posed, even small variations in Φ (in, e.g., the supremum norm on $\mathbb{R}^k$) may give rise to extremely large fluctuations in the solution f. It is clear from (7.10), however, that the mapping from Φ to f is continuous in the topology of uniform convergence on compacts on C^k.

Example 7.5. *The Wiener model.* Let us consider Example 3.1 from the point of view of reversed martingales. The model is given by (3.1), and, since the representation (7.2) in reverse time does not depend on the choice of P, we may as well consider the simplest case $a = 0$, i.e., we want to reverse the process

$$\begin{aligned} &dX_t = dW_t, \\ &X_0 = 0. \end{aligned} \tag{7.12}$$

It follows easily from [6] that

$$d\tilde{X}_r = -\frac{\tilde{X}_r}{T-r}dr + d\tilde{W}_r, \tag{7.13}$$

in other words

$$\begin{aligned} &\tilde{\mu}(r, x) = -\frac{x}{T-r}, \\ &\tilde{\sigma}(r, x) \equiv 1, \\ &\bar{\mu}(t, x) = -\frac{x}{t}, \\ &\bar{\sigma}(t, x) \equiv 1. \end{aligned} \tag{7.14}$$

The boundary value problem (7.5) now becomes

$$\begin{aligned} &\frac{\partial f}{\partial t} + \frac{x}{t}\frac{\partial f}{\partial x} - \frac{1}{2}\frac{\partial^2 f}{\partial x^2} = 0, \\ &f(T, x) = \Phi(x). \end{aligned} \tag{7.15}$$

Thus we are basically inverting the heat equation.

The complex valued stochastic differential equation (7.8) becomes

$$\begin{aligned} &\mathrm{d}Y_s = \frac{Y_s}{s}\mathrm{d}s + \mathrm{i}\,\mathrm{d}W_s, \\ &Y_t = x. \end{aligned} \tag{7.16}$$

and this equation can easily be solved as

$$Y_T = \frac{T}{t}x + \mathrm{i}\int_t^T \frac{T}{s}\mathrm{d}W_s. \tag{7.17}$$

The stochastic integral in (7.17) is Gaussian with zero mean and variance $D^2(t, T)$ where

$$D^2(t, T) = (T-t)T/t, \tag{7.18}$$

so the stochastic representation formula for the solution of (7.15) is

$$f(t, x) = \frac{1}{\sqrt{2\pi D^2}}\int_{-\infty}^{\infty} \Phi\left(\frac{T}{t}x + \mathrm{i}y\right)\exp\left\{-\frac{y^2}{2D^2}\right\}\mathrm{d}y. \tag{7.19}$$

From Theorem 7.3 we know that *if* (7.15) has a solution then the solution is given by (7.19). Now we turn the argument around and use the formal expression (7.19) to *define f*, which then can be seen to solve (7.15). Proceeding this way we may give a rather complete characterization of the optimal predictors for the Wiener model. First, however, we define a fairly wide class of functions within which we will search for solutions to (7.15).

Definition 7.6. A function

$$f: [d, T]\times\mathbb{R}\to\mathbb{R}$$

is said to belong to the class $E[d, T]$ if for all t in $[d, T]$ we have

$$f(t, x) \leq A\exp\{x^2/(2c^2)\}, \tag{7.20}$$

where A, c may depend on t and we demand that

$$c^2(t) > t. \tag{7.21}$$

Thus $f \in E[d, T]$ will imply that for all t in $[d, T]$, $f(t, X_t)$ is integrable.
Now we can give the central theorem for the Wiener model.

Theorem 7.7. *Equation* (7.15) *has a solution defined on* $(d, T]\times\mathbb{R}$ *with* $f \in [d, T]$ *if and only if* Φ *satisfies the following conditions*:

(i) Φ *is an entire analytic function on* C,

(ii) $$|\Phi(x+\mathrm{i}y)| \leq B\exp\left\{\frac{x^2}{2\alpha^2}\right\}\exp\left\{\frac{y^2}{2\beta^2}\right\},$$

where B is an arbitrary constant and α, β satisfies

(iii) $\alpha^2 \geq T, \qquad \beta^2 \geq \dfrac{T(T-d)}{d}.$

Furthermore, the solution is given by the formula (7.19).

Proof. Suppose first that $f \in E[d, T]$ in fact solves (7.15) on $(d, T]$ and that f is continuous on $[d, T]$. Then, using $f(d, \cdot)$ as boundary condition for $t = d$, we can easily solve the PDE in (7.15). In particular for $t = T$ we obtain

$$\Phi(x) = \int_{-\infty}^{\infty} f(d, y) \exp\left\{-\frac{1}{2b^2}(y - kx)^2\right\} \mathrm{d}y, \tag{7.22}$$

where

$$b^2 = d(T-d)/T, \qquad k = d/T. \tag{7.23}$$

From (7.22) and the fact that $f \in E[d, T]$ it follows immediately that Φ can be extended to an entire analytic function and an easy calculation shows that (ii)-(iii) are satisfied with equality holding in (iii).

Suppose on the other hand that Φ satisfies (i)-(ii) for some constants α and β. Then we may define f on $(d, T]$ by (7.19), where d is given by

$$d = T^2/(T + \beta^2). \tag{7.24}$$

It now follows from (ii) that we may change the path of integration in (7.19) to obtain

$$f(t, x) = \frac{1}{\sqrt{2\pi D^2}} \int_{-\infty}^{\infty} \Phi(\mathrm{i}y) \exp\left\{\frac{1}{2D^2}\left(\mathrm{i}y - \frac{T}{t}x\right)^2\right\} \mathrm{d}y. \tag{7.25}$$

Again using (ii) we see that we are permitted to differentiate under the integral sign, and it is easy to check that f actually solves (7.15). It is also obvious from (7.19) that we have the correct boundary values $f(T, x) = \Phi(x)$. □

Let us for illustrative purposes compute the optimal predictor for the Wiener model in the particular case when $\Phi(x) = x^n$. From (7.19) we see that

$$f(t, x) = \sum_{k=0}^{n} \binom{n}{k} \cdot \mathrm{i}^k \cdot \left(\frac{T}{t}\right)^{n-k} x^{n-k} \cdot E[Z^k], \tag{7.26}$$

where Z is normally distributed with zero mean and variance D^2. It is easily seen that, because Φ is real-valued, the imaginary parts in (7.26) must cancel. Using standard properties of the normal distribution we thus have

$$f(t, x) = \sum_{2k \leq n} \binom{n}{2k} (-1)^k \left(\frac{T}{t}\right)^{n-k} (T-t)^k m_k x^{n-2k}, \tag{7.27}$$

where

$$m_k = \prod_{r=1}^{k} (2r-1).$$

Example 7.8. Geometric Brownian motion.
For the model (3.10) we note that

$$X_t = \exp[(a - \tfrac{1}{2})t + W_t], \tag{7.28}$$

so we see that the task of predicting $\Phi(X_T)$ for the geometric Brownian model (3.10) is equivalent to that of predicting $\Phi_0(Y_T)$ for the Wiener model (3.1), where $\Phi_0(y) = \Phi[\exp(y)]$ and $Y_t = \log X_t$. We may thus apply Theorem 7.7 to the function Φ_0 and we have the following result.

Proposition 7.9. *Consider the geometric Brownian motion model and a given Φ. Suppose furthermore that Φ_0 satisfies the conditions of Theorem 7.7 where Φ_0 is defined as above. Then an optimal predictor f exists and is given by*

$$f(t, x) = \frac{1}{\sqrt{2\pi D^2}} \int_{-\infty}^{\infty} \Phi(x^{T/t}\, e^{iy}) \exp\left\{-\frac{y^2}{2D^2}\right\} dy. \tag{7.29}$$

Proof. From the argument above it is obvious that $f(t, x) = f_0(t, \log x)$, where f_0 is the optimal predictor for Φ_0 (in a Wiener model). Theorem 7.7 does the rest. □

To exemplify we see from (7.29) that for the choice $\Phi(x) = x$ we obtain the predictor

$$f(t, x) = x^{T/t} \exp(\tfrac{1}{2} D(t, T)^2). \tag{7.30}$$

8. Reversed point processes

In this section we look at some point process examples from the point of view of reverse martingales. In particular we show why the predictor (5.13) turns up in connection with the maximal family generated by the Yule model.

Example 8.1. *The Poisson model.* With N as in Example 3.4 we set

$$\tilde{N}_r = N_{T-r}^- = N_{(T-r)-}, \quad 0 \leq r \leq T.$$

For any P in the Poisson model (or in fact any P in the maximal family, see Example 5.4) the infinitesimal operator $\tilde{A}$ of $\tilde{N}$ is given by

$$[\tilde{A}(r)\tilde{f}(r, \cdot\,)](m) = \tilde{\lambda}(r, m)\{\tilde{f}(r, m-1) - \tilde{f}(r, m)\}, \tag{8.1}$$

where

$$\tilde{\lambda}(r, m) = m/(T - r), \quad 0 \leq r \leq T.$$

It is now an easy exercise to check that the predictor $f(t, n)$ in (4.5) satisfies (6.12).

Example 8.2. *The Yule model.* We consider the Yule model of Example 3.5 and the corresponding maximal family discussed in Example 5.5.

As mentioned in Example 5.5 no solution to the basic equation exists for most interesting choices of the function Φ. The viewpoint taken in this section hopefully gives more insight into what goes wrong. Let

$$\tilde{N}_r = N^-_{T-r}, \quad \tilde{Y}_r = Y^-_{T-r}, \qquad 0 \leqslant r \leqslant T,$$

With (N, Y) as in Example 5.5. It can be shown (Johansson [9]) that, for each P in the maximal family, the process

$$\tilde{N}_r + \int_0^r \tilde{\lambda}(u, \tilde{N}_u, \tilde{Y}_u)\, du + I\{r > T-\tau\}, \quad 0 \leqslant r \leqslant T, \tag{8.2}$$

is an $\tilde{\mathcal{F}}^{(N,Y)}$-martingale, where

$$\tilde{\lambda}(r, m, x) = \frac{g_{m-1}(x/(T-r)-1)}{g_m(x/(T-r))} \cdot \frac{m}{T-r}, \tag{8.3}$$

$$\tau = \inf\{t \geqslant 0;\ N_t = 1\},$$

and g_n is as in Example 5.5. The indicator function appearing in (8.2) has to do with the fact that the last jump of $\tilde{N}$ (which is the first jump of N) is deterministically determined by the value of $\tilde{Y}$ (which is the integral of N). Now, when looking for a function f which will make the process $f(t, N_t, Y_t)$ a reverse martingale, it is natural to apply the change of variable formula to $F = \tilde{f}$.

If we do this formally without bothering about technical details we obtain

$$\begin{aligned} F(r, \tilde{N}_r, \tilde{Y}_r) &= F(0, \tilde{N}_0, \tilde{Y}_0) \\ &\quad - \int_{(0,r]} [F(u, \tilde{N}^-_u - 1, \tilde{Y}^-_u - T + u) - F(u, \tilde{N}^-_u, \tilde{Y}^-_u)] \\ &\qquad \times [d\tilde{N}_u + \tilde{\lambda}(u, \tilde{N}_u, \tilde{Y}_u)\, du] \\ &\quad + \int_{(0,r]} \Big[\frac{\partial}{\partial u} F(u, \tilde{N}_u, \tilde{Y}_u) \\ &\qquad + \{F(u, \tilde{N}_u - 1, \tilde{Y}_u - T + u) - F(u, \tilde{N}_u, \tilde{Y}_u)\} \\ &\qquad \times \tilde{\lambda}(u, \tilde{N}_u, \tilde{Y}_u) \Big]\, du. \end{aligned} \tag{8.4}$$

If we simply disregard the indicator process in (8.2), which is a minor nuisance, it is natural to look for solutions to the equation

$$\begin{aligned} &\frac{\partial}{\partial u} F(u, m, x) + [F(u, m-1, x-T+u) - F(u, m, x)]\tilde{\lambda}(u, m, x) = 0, \\ &F(0, m, x) = \Phi(m, x), \end{aligned} \tag{8.5}$$

and it can be shown that $F(r, m, x) = f(T-r, m, x)$ with f as in (5.13) satisfies (8.5). The predictor (5.13) is thus somewhat carelessly derived, but in fact it comes very close to being unbiased. For more of the (rather forbidding) difficulties which surround the basic equation for this model see [9].

For the concrete choices $\Phi(n, y) = n$ and $\Phi(n, y) = y$ the predictor (5.13) reduces to

$$f(t, N_t, Y_t) = N_t + \int_t^T \hat{\lambda}_u(t, N_t, Y_t)\, du, \tag{8.6}$$

and

$$f(t, N_t, Y_t) = N_t + \int_t^T u\hat{\lambda}_u(t, N_t, Y_t)\, du, \tag{8.7}$$

respectively, where

$$\hat{\lambda}_u(t, m, x) = \frac{g_{m-1}((x-u)/t)}{g_m(x/t)} \cdot \frac{m}{t}. \tag{8.8}$$

9. Prediction errors and information inequalities

We now go back to the setting of Section 6 and study the mean square error

$$C(t, T, \Phi, P) = E_P[\{f(t, X_t) - \Phi(X_T)\}^2]. \tag{9.1}$$

From an abstract point of view it is easy enough to obtain an expression for C. Let M be defined as in (6.2) and define $\tilde{M}$ by

$$\tilde{M}_r = M_{T-r} = f(T-r, X_{T-r}). \tag{9.2}$$

Suppose now that f is an UMSEUP for $\Phi(X_T)$. Then, by Theorem 6.2, $\tilde{M}$ is a reverse martingale, so if we denote its quadratic variation process by $\langle \tilde{M} \rangle$ we have the following result.

Proposition 9.1. *Suppose that f is an UMSEUP. Then, suppressing T and Φ:*

$$C(t, P) = E_P[\langle \tilde{M} \rangle_{T-t} - \langle \tilde{M} \rangle_0]. \tag{9.3}$$

If $\tilde{X}$ is a diffusion of the form (7.2) *then*

$$C(t, P) = \int_t^T E_P[\|\nabla_x f(s, X_s)\bar{\sigma}(s, X_s)\|^2]\, ds, \tag{9.4}$$

where $\bar{\sigma}$ is given by (7.4).

Proof. If $\tilde{X}$ is a diffusion and f is an UMSEUP then, by Theorem 6.2 and Itô's formula

$$d\tilde{M}_r = (\nabla_x \tilde{f})(s, \tilde{X}_r)\tilde{\sigma}(s, \tilde{X}_r)\, d\tilde{W}_r, \tag{9.5}$$

which gives us (9.4). □

Even in a simple concrete case it may be very hard to compute f so the need arises for a simpler estimate for C. In this section we will therefore derive some information inequalities of Cramér-Rao type. The present discussion will be somewhat informal, but in Sections 10-11 we will give precise results. First we recall Definition 5.1 which we state again for easier reference.

Definition 9.2. Consider T as fixed. For any y in $\mathbb{R}^k$ we define the probability measure

$$P^{T,y} \text{ on } (\Omega, \mathcal{F}_T^X)$$

by

$$P^{T,y}(A) = P(A \mid X_T = y).$$

We define expectation operators $E^{T,y}$ analogously.

By predictive sufficiency this definition is independent of the choice of P, and since T is fixed we often suppress it in the sequel.

The basic equation (4.1) now reads

$$E^y[f(t, X_t)] = \Phi(y), \quad \mathcal{P} \circ X^{-1}\text{-a.s.} \tag{9.6}$$

so we see that f is an optimal predictor if and only if it is an unbiased parameter estimate of $\Phi(y)$, where y acts as the parameter. Consequently we may now apply standard Cramér-Rao reasoning to (9.6).

Let us therefore denote the restriction of $P^{T,y}$ to $\mathcal{F}_t^X$ by $P_t^{T,y}$, and let us assume that

$$P_t^{T,y} \ll m, \tag{9.7}$$

for some base measure m.

Defining a family of Radon-Nikodym derivatives by

$$L_t^{T,y} = \frac{\mathrm{d}P_t^{T,y}}{\mathrm{d}m}, \tag{9.8}$$

and assuming enough regularity we have the following standard Cramér-Rao inequality (see Lehmann [13] for details).

$$E^y[\{f(t, X_t - \Phi(X_T)\}^2] \geq \nabla\Phi(y)^* I(t, y)^{-1} \nabla\Phi(y), \tag{9.9}$$

where the gradient is regarded as a column vector, * denotes transpose and the information matrix $I(t, y)$ is given by

$$I_{i,j}(t, y) = E^{T,y}\left[\left(\frac{\partial}{\partial y_i} \log L_t^{T,y}\right)\left(\frac{\partial}{\partial y_j} \log L_t^{T,y}\right)\right] = -E^{T,y}\left[\frac{\partial^2}{\partial y_i\, \partial y_j} \log L_t^{T,y}\right]. \tag{9.10}$$

Integrating over y we now obtain our main Cramér-Rao inequality.

$$E_P[\{f(t, X_t) - \Phi(X_T)\}^2] \geq E_P[\nabla\Phi(X_T) * I(t, X_T)^{-1} \nabla\Phi(X_T)]. \tag{9.11}$$

In the scalar case $k = 1$ we can write (9.11) as

$$E_P[\{f(t, X_t) - \Phi(X_T)\}^2] \geq E_P\left[\frac{\{\Phi'(X_T)\}^2}{I(t, X_T)}\right], \tag{9.12}$$

where

$$I(t, y) = E^{T,y}\left[\left\{\frac{\partial}{\partial y} \log L_t^{T,y}\right\}^2\right]. \tag{9.13}$$

The next two sections are devoted to the explicit calculation of $I(t, y)$ and the right-hand side of (9.11) for some concrete classes of models.

10. Cramér–Rao inequalities for diffusion models

For the moment we forget about the family of probability measures $\mathcal{P}$, and just consider a k-dimensional diffusion process defined on a fixed probability space

$$\begin{aligned} &dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dW_t, \\ &X_0 = x_0. \end{aligned} \tag{10.1}$$

We assume that X has a smooth strictly positive transition density function $q(t, x; s, y)$ for $t < s < T$, where T is fixed.

Definition 10.1. For each y in $\mathbb{R}^k$ we define the process $b(t, X_t, y)$ by

$$b(t, X_t, y) = \sigma^*(t, X_t)\nabla_x \log q(t, X_t; T, y), \tag{10.2}$$

where * denotes transpose and the gradient is regarded as a column vector. We will often suppress y. We also define the matrix processes α and H by

$$\alpha(t) = \sigma(t, X_t)\sigma^*(t, X_t), \tag{10.3}$$

$$H_{i,j}(t, y) = \frac{\partial^2}{\partial x_i\,\partial y_j} \log q(t, X_t; T, y). \tag{10.4}$$

Definition 10.2. We say that X is CR-regular if, for all y, the following conditions hold for all $t < T$.

(i) $$E\left[\int_0^t \|b(s, X_s, y)e\|^2\,ds\right] < \infty,$$

(ii) $$E\left[\int_0^t \left\|\frac{\partial^2}{\partial y_i\,\partial y_j} b(s, X_s, y)\right\|^2 ds\right] < \infty,$$

(iii) $$\frac{\partial^2}{\partial y_i\,\partial y_j}\int_0^t b(s, X_s, y)\,dW_s = \int_0^t \frac{\partial^2}{\partial y_i\,\partial y_j} b(s, X_s, y)\,dW_s,$$

(iv) $$\frac{\partial^2}{\partial y_i\,\partial y_j}\int_0^t \|b(s, X_s, y)\|^2\,ds = 2\int_0^t b_i^*(s, X_s, y)b_j(s, X_s, y)\,ds + 2\int_0^t b_{i,j}^*(s, X_s, y)b(s, X_s, y)\,ds,$$

where the subscripts i,j denote partial derivatives with respect to y.

We now have the following basic result.

Proposition 10.3. *Suppose that f is such that* (9.6) *holds and suppose that X is CR-regular. Then we have the Cramér–Rao inequality* (9.9), *where the information matrix $I(t, y)$ is given by*

$$I(t, y) = E^{T,y}\left[\int_0^t H^*(s, y)\alpha(s)H(s, y)\,\mathrm{d}s\right]. \tag{10.5}$$

Proof. When we consider X under the measure $P^{T,y}$ we are 'pinning' X at $X_T = y$. Thus we have a reciprocal process as developed in Jamison [7], and it follows from CR-regularity and Jamisons results that, under $P^{T,y}$, X will satisfy the following stochastic differential equation for $0 \leq t < T$.

$$\begin{aligned} \mathrm{d}X_t &= \{\mu(t, X_t) + \sigma(t, X_t)b(t, X_t)\}\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}V_t, \\ X_0 &= x_0, \end{aligned} \tag{10.6}$$

where V is a $P^{T,y}$-Wiener process given by

$$\mathrm{d}V_t = \mathrm{d}W_t - b(t, X_t)\,\mathrm{d}t. \tag{10.7}$$

In other words we have $P^{T,y} \ll P$ on $\mathcal{F}_t^X$ for $t < T$, and the Girsanov Theorem gives us

$$\log L_t^{T,y} = \int_0^t b^*(s, X_s)\,\mathrm{d}W_s - \tfrac{1}{2}\int_0^1 b^*(s, X_s)b(s, X_s)\,\mathrm{d}s. \tag{10.8}$$

Using the regularity assumptions to differentiate (10.8) we easily obtain

$$\frac{\partial^2}{\partial y_i\,\partial y_j}\log L_t^{T,y} = \int_0^t b_{i,j}^*\,\mathrm{d}W_s - \int_0^t b_i^* b_j\,\mathrm{d}s - \int_0^t b_{i,j}^* b\,\mathrm{d}s. \tag{10.9}$$

We substitute (10.7) into the stochastic integral in (10.9), which gives us

$$\frac{\partial^2}{\partial y_i\,\partial y_j}\log L_t^{T,y} = \int_0^t b_{i,j}^*\,\mathrm{d}V_s - \int_0^t b_i^* b_j\,\mathrm{d}s. \tag{10.10}$$

Now we take $E^{T,y}$-expectations in (10.10) and use the fact that V is a $P^{T,y}$-Wiener process. This, together with (9.10), gives us

$$I(t, y)_{i,j} = E^{T,y}\left[\int_0^t b_i^*(s, X_s)b_j(s, X_s)\,\mathrm{d}s\right], \tag{10.11}$$

which gives us (10.5). □

Now we can go back to the prediction setting of Section 4, i.e., we consider a model $(\Omega, \mathcal{F}, \mathcal{P}, X)$ where X is supposed to be prediction sufficient.

Theorem 10.4. *Suppose that $f(t, X_t)$ is an unbiased estimator of $\Phi(X_T)$ in the sense that f solves the basic equation* (4.1). *Suppose furthermore that X is CR-regular for every P in $\mathcal{P}$. Then, for every P in $\mathcal{P}$, we have the Cramér–Rao inequality* (9.11) *where the information matrix $I(t, y)$ is given by* (10.5).

Proof. The result follows immediately from Proposition 10.3 and the fact that, by predictive sufficiency, every P in $\mathcal{P}$ will generate the same $P^{T,y}$. □

We note that if X is CR-regular for every P in the maximal family $\mathcal{M}$, then (9.11) will in fact hold for all P in $\mathcal{M}$.

Corollary 10.5. *For the Wiener model* (3.1) *we have, for any unbiased* $\mathcal{F}_t^X$*-predictor* Z:

$$E_P[\{Z-\Phi(X_T)\}^2] \geq \frac{T(T-t)}{t} E_P[\{\Phi'(X_T)\}^2], \tag{10.12}$$

for all P *in* $\mathcal{M}$.

Proof. When we compute q we may as well (because of prediction sufficiency) set $a=0$. Then we have

$$q(s,x;T,y)=\frac{1}{\sqrt{2\pi(T-s)}}\exp\left\{-\frac{(y-x)^2}{2(T-s)}\right\}, \tag{10.13}$$

so from (10.5) we have

$$I(T,y)=\int_0^t \frac{1}{(T-s)^2}\,ds=\frac{t}{T(T-t)}. \qquad \square \tag{10.14}$$

We note that the inequality is sharp in the sense that equality holds when $\Phi(y)=y$, and Z is the predictor (4.24).

11. An information inequality for the Poisson model

For the Poisson model (Example 3.4) we cannot use the inequality (9.11) since, in this case, the parameter y will be integer valued. We will instead use the following result, which will be referred to as the HCR-inequality (Hammersley-Chapman-Robbins, see Lehmann [13]).

Proposition 11.1. *Let* $(\Omega, \mathcal{F})$ *be a measurable space and let* $\{P^n: n=0,1,2,\dots\}$ *be a family of probability measures on* $(\Omega, \mathcal{F})$ *such that*

$$P^{n-1} \ll P^n, \tag{11.1}$$

on $(\Omega, \mathcal{F})$ *for all* $n \geq 1$.

Denote the corresponding expectation operators by E^n, *and suppose that* Z *is an* $\mathcal{F}$*-measurable random variable such that*

$$E^n[Z^2] < \infty, \quad n \geq 1, \tag{11.2}$$

and satisfying

$$E^n[Z] = \Phi(n), \quad n \geq 1. \tag{11.3}$$

for some real valued function Φ.

Then we have

$$E^n[\{Z-\Phi(n)\}^2] \geq \frac{\{\Phi(n)-\Phi(n-1)\}^2}{E^n[\{L^n-1\}^2]}, \quad n \geq 1, \tag{11.4}$$

where

$$L^n = \frac{dP^{n-1}}{dP^n}. \tag{11.5}$$

Proof. By the Schwartz inequality

$$(E^n[\{Z-\Phi(n)\}\{L^n-1\}])^2 \leq E^n[\{Z-\Phi(n)\}^2] \cdot E^n[\{L^2-1\}^2]. \tag{11.6}$$

Furthermore, for $n \geq 1$ we have

$$\begin{aligned} &E^n[\{Z-\Phi(n)\}\{L^n-1\}] \\ &= E^n[\{Z-\Phi(n)\}L^n] \\ &= E^n[\{Z-\Phi(n-1)\}L^n] + E^n[\{\Phi(n-1)-\Phi(n)\}L^n] \\ &= E^{n-1}[Z-\Phi(n-1)] + \{\Phi(n-1)-\Phi(n)\}E^n[L^n] \\ &= \Phi(n-1)-\Phi(n). \end{aligned}$$

Which, with (11.6), yields the result. □

Now we consider the Poisson model of Example 3.4. We define $P_t^{T,n}$ on $\mathcal{F}_t^N$ (for $t \leq T$) by

$$P_t^{T,n}(Z) = P_a(Z \mid N_T = N), \quad m \geq 0, \tag{11.7}$$

where the definition does not depend on the choice of the parameter a. It is now easily seen that

$$P_t^{T,n} \ll P_a \quad \text{on } \mathcal{F}_t^N,$$

for all $n \geq 0$, and that

$$\frac{dP_t^n}{dP_a} = \frac{p_{t,T}^a(N_t; n)}{p_{o,T}^a(0; n)} \quad \text{on } \mathcal{F}_t^N, \tag{11.8}$$

where

$$p_{s,t}^a(m; n) = P_a(N_t = n \mid N_s = m).$$

Furthermore $P_t^{n-1} \ll P_t^n$ on $\mathcal{F}_t^N$ for $t < T$, so we have

$$L_t^{T,n} = \frac{dP_t^{T,n-1}}{dP_t^{T,n}} = \frac{dP_t^{T,n-1}/dP_a}{dP_t^{T,n}/dP_a} = \frac{n-N_t}{n} \cdot \frac{T}{T-t}. \tag{11.9}$$

Proposition 11.2. *Fix t, T and Φ with $t < T$. Suppose that Z is any unbiased $\mathcal{F}_t^N$-predictor of $\Phi(N_T)$, i.e., Z solves*

$$E_t^{T,n}[Z] = \Phi(n), \quad n \geq 0. \tag{11.10}$$

Then we have, for all $a > 0$:

$$E_a[\{Z-\Phi(N_T)\}^2] \geq \frac{T-t}{t} E_a[N_T\{\Phi(N_T)-\Phi(N_T-1)\}^2]. \tag{11.11}$$

Proof. It is easily seen that, under $P^{T,n}$, N_t has a Bin$(n, t/T)$-distribution, so using (11.9) we have

$$E^{T,n}[\{L_t^{T,n}-1\}^2] = E^{T,n}\left[\left\{\frac{n-N_t}{n}\cdot\frac{T}{T-t}-1\right\}^2\right]$$
$$= \left\{\frac{T}{n(T-t)}\right\}^2 E^{T,n}\left[\left\{N_t - n\frac{1}{T}\right\}^2\right]$$
$$= \left\{\frac{T}{n(T-t)}\right\}^2 n\frac{t}{T}\cdot\frac{T-t}{T} = \frac{t}{n(T-t)}.$$

Using the HCR-inequality (11.4) we thus have

$$E^{T,n}[\{Z-\Phi(N_T)\}^2] \geq \frac{n(T-t)}{t}\{\Phi(n)-\Phi(n-1)\}^2, \tag{11.12}$$

and taking expectations in (9.12) we obtain (9.11). □

We remark that the result is sharp in the sense that equality holds when $\Phi(n) = n$, and Z is the predictor (4.7).

Acknowledgement

The authors would like to thank an anonymous referee for a number of helpful comments. We would also like to thank Per Appelgren, M.Sc. for his kind assistance with the 'Math Type'-system.'

References

[1] A.T. Bharucha-reid, Elements of the Theory of Markov Processes and their Applications (McGraw-Hill, New York, 1960).
[2] J.F. Bjørnstad, Predictive likelihood: A review, Statist. Sci. 5 (1990) 242-265.
[3] P. Brémaud, Point Processses and Queues (Springer, Berlin, 1981).
[4] C. Dellacherie and P.-A. Meyer, Probabilities and Potential B (North-Holland, Amsterdam, 1982).
[5] S. Giesser, On the prediction of observables: a selective update, Bayesian Statist. 2 (1985) 203-230.
[6] U.G. Haussmann and E. Pardoux, Time reversals of diffusions, Ann. Probab. 14 (1986) 1188-1205.
[7] B. Jamison The Markov processes of Schrödinger, Z. Wahrsch. Verw. Gebiete 32 (1975) 323-331.
[8] J.L. Jensen and B. Johansson, The extremal family generated by the Yule process, Stochastic Process. Appl. 36 (1990) 59-76.
[9] B. Johansson, Predictive inference and extremal families, Tech. Rept. TRITA-MAT-1989-2, The Royal Inst. of Technology (Stockholm, 1989).
[10] B. Johansson, Unbiased prediction in the Poisson and Yule processes, Scand. J. Statist. 17 (1990) 135-145.
[11] N. Keiding, Estimation in the birth process, Biometrica 61 (1974) 71-80.
[12] S.L. Lauritzen, Extremal families and systems of sufficient statistics, Lecture Notes in Statist. No. 49 (Springer, Berlin, 1988).
[13] E.L. Lehmann, Testing Statistical Hypotheses (Wiley, New York, 1986).
[14] K. Takeushi and M. Akahira, Characterization of prediction sufficiency (adequacy) in terms of risk functions, Ann. Statist. 3 (1975) 1018-1024.